

Im2Text: Describing Images Using 1 Million Captioned Photographs

Vicente Ordonez (presenter), Girish Kulkarni, Tamara L. Berg
Stony Brook University



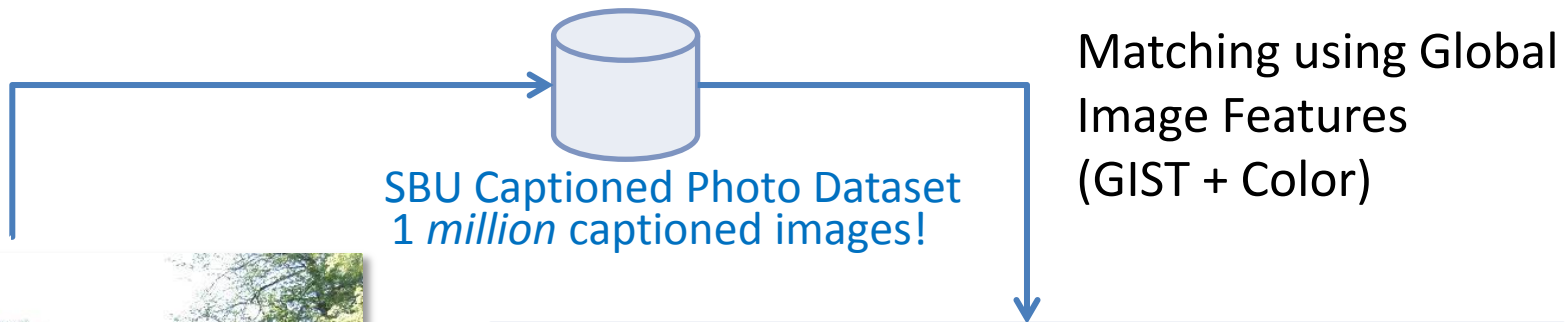
Computer Vision

sky
trees
water
building
bridge

Our Goal

An old bridge over dirty green water.
One of the many stone bridges in town
that carry the gravel carriage roads.
A stone bridge over a peaceful river.

Harness the Web!



SBU Captioned Photo Dataset
1 million captioned images!

Matching using Global Image Features
(GIST + Color)



Smallest house in paris
between red (on right)
and beige (on left).



Bridge to temple in
Hoan Kiem lake.



A walk around the
lake near our house
with Abby.



The water is clear
enough to see
fish swimming
around in it.



Hangzhou bridge in
West lake.

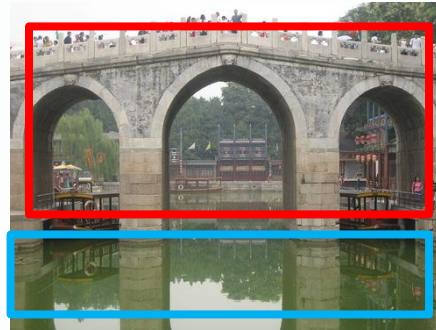


The daintree river by
boat.

Transfer Caption(s)

e.g. "The water is clear
enough to see fish
swimming around in it."

Use High Level Content to Rerank (Objects, Stuff, People, Scenes, Captions)



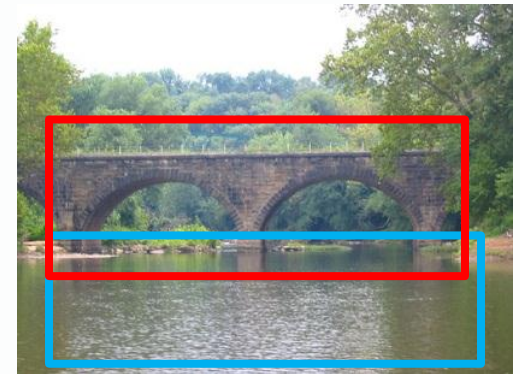
The **bridge over** the lake on Suzhou Street.



Iron **bridge over** the Duck **river**.



The Daintree **river** by boat.



Bridge over Cacapon **river**.

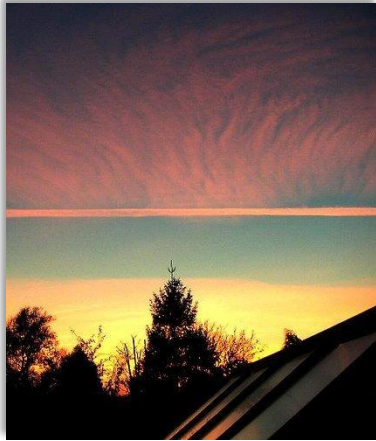
Transfer Caption(s) ←

e.g. “The bridge over the lake on Suzhou Street.”

• • •

Results

Good



Amazing colours in the sky at sunset with the orange of the cloud and the blue of the sky behind.



Cat in sink.



A female Mallard duck in the lake at Luukki Espoo.



Fresh fruit and vegetables at the market in Port Louis Mauritius.

Bad



The cat in the window.



The boat ended up a kilometre from the water in the middle of the airstrip.